基于深度残差反投影注意力网络的图像超分辨率 *

胡高鹏[a]，陈子鎏[a]，王晓明[a, b†]，张开放[a]，黄增喜[a]，杜亚军[a]

(西华大学 a. 计算机与软件工程学院; b. 机器人研究中心，成都 610039)

摘 要：针对多数单帧图像超分辨率(single image super-resolution，SISR)重建方法存在的特征信息发掘不充分、特征图各通道之间的相互依赖关系难以确定以及重建高分辨率(high resolution，HR)图像时存在重构误差等问题，提出了基于深度残差反投影注意力网络的图像超分辨率(SR)算法。即利用残差学习的思想缓解训练难度和充分发掘图像的特征信息，并使用反投影学习机制学习高低分辨图像之间的相互依赖关系，此外引入了注意力机制动态分配各特征图以不同的注意力资源从而发掘更多的高频信息和学习特征图各通道之间的依赖关系。实验结果表明了所提方法相比于多数单帧图像超分辨率方法，不仅在客观指标方面得到了显著的提升，而且重建的预测图像也具有更加丰富的纹理信息。

关键词：注意力机制；超分辨率；反投影；残差学习；卷积神经网络

中图分类号：TP391.4 doi: 10.19734/j.issn.1001-3695.2020.03.0081

Image super-resolution based on depth residual back projection attention network

Hu Gaopeng[a], Chen Ziliu[a], Wang Xiaoming[a, b†], Zhang Kaifang[a], Huang Zengxi[a], Du Yajun[a]

(*a. School of Computer & Software Engineer*, *b. Robotics Research Center*, *Xihua University*, *Chengdu* 610039, *China*)

Abstract: Focused on the partly issue that in the process of single-frame image super-resolution reconstruction, such as insufficient utilization of feature information during image super-resolution reconstruction, the interdependence between the channels of the feature map is difficult to determine, and reconstruction errors existing at high-resolution image reconstructed, this paper proposed an single image super resolution methods based on depth residual backprojection attention network. It used the residual learning to ease the training difficulty and fully discover the feature information of the image, and used the back-projection method to learn the interdependence between the high- and low-resolution images. In addition, it introduced the attention mechanism to assign each feature map with different attention to discover more high-frequency information, and learnt the interdependence between the channels of the feature map. The experimental results show that compared with most single-frame image super-resolution methods, the proposed method not only has a significant improvement in objective indicators, but also the reconstructed predicted image has richer texture information.

Key words: attention mechanism; super-resolution(SR); back projection; residual learning; convolutional neural network

0 引言

单幅图像超分辨率旨在利用单张(low resolution, LR)图像重建 HR 图像。其在现实生活中得到了广泛的应用，如利用 SR 技术提高监控视频中人脸的识别精度，在 HDTV 中产生更好质量的视频和获得较高分辨率的医学图像等。经过不断的探索和发展，目前已产生了大量基于机器学习和深度学习的 SR 方法。鉴于，基于深度学习方法的明显优势，本文主要研究基于深度学习的 SISR 任务。

目前基于深度学习的 SR 算法按上采样的方式可以分为前端上采样方法[1~5]、后端上采样方法[6~9]、渐进式上采样方法[10]和迭代式上采样方法[11]等。Dong 等人[1]首次将卷积神经网络应用于 SR 任务，提出了(super-resolution convolutional neural network, SRCNN)方法，即预先使用双三次插值的方法将 LR 图像上采样至给定的倍数，然后使用三层的卷积神经网络学习其到 HR 图像之间的映射关系。随后 Kim 等人[2]提出的(super-resolution using very deep convolutional network, VDSR)引入残差的思想缓解了梯度消失或梯度爆炸的问题，将网络结构加深至 20 层，同时使用较大的感受野提取 LR 图像中的特征信息，不仅加快了收敛速度同时提升重建效果。Tai 等人[6]提出的(deep recursive residual network, DRRN)使用局部残差和全局残差的策略并引入递归的思想将网络加深至 52 层，并在增加深度的同时保持参数规模，使得重建效果得到了进一步的提升。Kim[7]等人提出的(deep recursive convolutional network, DRCN)使用递归的思想共享网络结构之间的参数，降低了训练的难度。Dong 等人[8]提出的(faster SRCNN, FSRCNN)，在卷积神经网络的后端使用反卷积的方法扩充图像的尺寸。 Shi 等人[9]提出的(efficient sub-pixel convolutional neural network, ESPCN)，使用亚像素卷积方法在网络结构的后端将学习到的 LR 特征映射至给定的分辨率。FSRCNN 和 ESPCN 等方法都证明了后端上采样策略有效的降低神经网络的计算复杂度，提高了 HR 图像空间分辨率。Lai 等人[10]提出的(Laplacian pyramid Networks, LapSRN)融合了拉普拉斯金字塔的思想，对于输入的 LR 图像渐进式学习图像的高频部分，低频部分只做双三次插值放大，这种策略不仅加快了学习进度同时提高了重建效果。

随后 Haris 等人[11]提出了(deep back-projection networks for super-resolution，DBPN)，其使用连续迭代的上采样和下

收稿日期：2020-03-09；**修回日期：**2020-04-25 **基金项目：**国家自然科学基金资助项目(61602390)；西华大学研究生创新基金资助项目(ycjj2019095)

作者简介：胡高鹏(1994-)，男，河南商丘人，硕士研究生，主要研究方向为模式识别、图像处理；陈子鎏(1996-)，男，广东韶关人，硕士研究生，主要研究方向为模式识别、图像处理；王晓明(1977-)，男(通信作者)，四川简阳人，副教授，硕导，博士，主要研究方向为模式识别、机器学习、图像处理、计算机视觉(wangxmxh@aliyun.com)；张开放(1994-)，男，河南周口人，硕士研究生，主要研究方向为模式识别、图像处理；黄增喜(1985-)，男，广西百色人，副教授，硕导，博士，主要研究方向为深度学习、生物特征识别、图像处理；杜亚军(1967-)，男，四川巴中人，教授，硕导，博士，主要研究方向为数据库理论、智能网络.

采样的策略，学习 HR 图像和 LR 图像之间的映射关系，并使用误差反馈机制纠正重构误差。虽然该方法使用了相互连接的上下采样策略和误差反馈机制学习 HR 和 LR 图像之间的映射关系，并获得了较好的重建效果。但是该方法使用了较深的网络结构以至于重建的 HR 图像较平滑，同时忽略了不同阶段产生的特征图对预测 HR 图像的贡献值存在差异性。针对以上这些问题本文提出了残差反投影注意力网络，即融合了残差思想和注意力机制，以缓解高频信息的丢失和学习特征图之间的相互依赖关系。本文的主要贡献有以下两点：

a) 融合了迭代反投影方法和残差思想，提出了残差反投影方法。该方法有效的降低了训练难度，并且减少了训练过程中信息的丢失，尽可能的保留了高频特征。

b) 本文引入并发展了注意力机制，提出了全局注意力单元，即对残差反投影块各阶段产生的特征图以及特征图的各通道，自动分配注意力资源。使得在重建 HR 图像时，可以发掘更多的高频信息。

1　相关工作

1.1　残差学习

在训练很深的网络结构时，由于初始化参数很接近于零，所以在网络反向转播更新参数时，容易导致梯度弥散。使得加深网络结构不仅不能提升网络性能，甚至使网络性能更差。针对此问题 He 等人[12]提出 ResNet，使用残差学习的思想缓解了梯度弥散的问题。其主要思想是假设设计的较深的网络结构，存在冗余层并且冗余层需要完成恒等映射 ，即保证输入输出完全相同。但是学习恒等映射 存在着一定的困难，ResNet 为了避免学习恒等映射的参数，使用了如图 1 所示的网络结构，即 $H(x)=F(x)+x$。其可转换 $F(x)=H(x)-x$，其中 $F(x)$ 为残差项，当残差项 $F(x)=0$ 时，就可以很容易的构造恒等映射 $H(x)=x$。相对于学习恒等映射 $H(x)=x$，则学习 $F(x)=0$ 更容易。

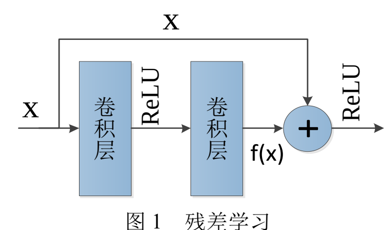


图 1　残差学习

Fig. 1　Residual learning

1.2　深度反投影网络

Haris 等人[11]提出了深度反投影网络 DBPN，其使用迭代的反投影方法来学习 LR 和 HR 图像之间的映射关系，并利用误差反馈机制纠正 LR 和 HR 图像之间的重构误差。DBPN 网络结构主要是由初始化特征提取单元、若干个相连的反投影单元和重建层组成，并引入密集连接策略将其发展为(Dense DBPN, D-DBPN)，其网络架构如图 2 所示。对于输入的 LR 图像，先进行初始化特征提取得到浅层特征 H_0，然后使用若干个迭代的上下反投影单元学习 HR 和 LR 特征之间的重构误差，最后级联先前各阶段产生的 HR 特征图并重建预测图像。其中每个反投影单元中都包含有上采样和下采样操作，其中上采样和下采样分别使用反卷积层和卷积层实现。

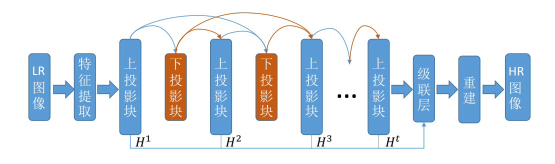


图 2 深度反投影网络

Fig. 2　Schematic diagram of deep back projection network

2　本文方法

虽然 D-DBPN[11]方法在各方面均已取得了令人满意的效果，但是仍然存在部分问题。如作者使用了较深的网络结构，造成学习过程中特征信息丢失和梯度弥散的问题。尽管作者使用了密集连接的方法对 DBPN 进行改进缓解了这些问题并进一步加深了网络结构，但是由于密集连接是在维度层面进行特征图级联，所以并未充分挖掘 HR 和 LR 空间的特征信息。同时其平等的对待各分层的特征信息，未考虑到各分层特征之间的依赖关系和各级特征之间的冗余信息。鉴于此本文提出了残差反投影注意力网络。

2.1　网络结构

本文所提出的残差反投影注意力网络结构如图 3 所示，其包括浅层特征提取单元、残差反投影单元(RBP)、全局注意力单元(GA)以及重建单元四个模块。其中在浅层特征提取单元中，本文使用两层的卷积层从输入的 LR 图像中提取得到浅层特征 。假设输入的 LR 图像和预测的 HR 图像分别为 和 ，浅层特征提取操作如式(1)所示。

$$F_0=\phi(\phi(I_{LR}))\tag{1}$$

其中 $\phi(\cdot)$ 表示对输入的 LR 图像 I_{LR} 的卷积操作，随后将得到的浅层特征 F_0 输入残差反投影单元，以发掘更深层次的高频特征信息。该过程如式(2)所示。

$$F_{DF}=H_{RBP}(F_0)\tag{2}$$

其中 F_{DF} 表示经过残差反投影单元得到的特征，$H_{RBP}(\cdot)$ 表示本文所提出的残差反投影单元结构，其包含 N 个由残差连接的反投影单元，并且每个反投影单元包括上投影块和下投影块。其利用相互连接的上投影块和下投影块学习各种类型上下采样算子之间的重构误差，然后使用误差反馈机制，纠正 HR 和 LR 特征信息之间的映射关系。每个投影单元都级联了之前各投影单元输出的 HR 特征信息，并且引入了残差思想，使用跳跃连接的方式来增强特征学习，具体如 3.2 节介绍。

随后，使用全局注意力机制单元学习残差反投影单元所级联在一起的各投影单元产生的 HR 特征图和特征图的各个通道的注意力资源分配大小，如式(3)所示。

$$F_{AT}=H_{GA}(F_{DF})\tag{3}$$

其中 F_{AT} 表示经过全局注意力单元分配注意力资源的特征信息，$H_{GA}(\cdot)$ 表示全局注意力单元。全局注意力单元依据其输入的特征信息对重建预测图像的贡献大小和各特征图之间的相互依赖关系，自动对各特征图以及特征图的各通道分配不同的注意力资源。最后对全局注意力单元产生的特征信息 F_{AT} 使用单层的卷积层重建，如式(4)所示。

$$I_{SR}=H_{REC}(F_{AT})=H_{RBPAN}(I_{LR})\tag{4}$$

其中 $H_{RBPAN}(\cdot)$ 和 $H_{REC}(\cdot)$ 分别代表残差反投影网络和重建单元，其中重建单元是由单层的卷积层组成。

此外，在 3.1 节详细分析了适合本文方法的损失函数，同时由于 L_1 范数本身的稀疏性和在文献[20]中被证明的较快的收敛性，故而选择 L_1 范数作为本文方法的优化目标。假设给定的训练集 $\{I_{LR}^i,I_{HR}^i\}_{i=1}^N$，其包含 N 个 LR 输入图像和对应的 HR 真实图像，则本文方法的优化目标是最小化 L_1 范数损失函数，损失函数定义如式(5)所示。

$$L(\theta)=\frac{1}{N}\sum_{i=1}^{N}\left\|H_{RBPAN}(I_{LR})-I_{HR1}^i\right\|_1\tag{5}$$

其中 θ 代表本文网络结构中所涉及的参数。此外本文使用随机梯度下降法最小化 L_1 损失函数。由于浅层特征提取单元和重建单元并无特殊之处，故随后本文将重点介绍残差反投影单元和全局注意力单元。

2.2　残差反投影单元

本节将详细介绍残差反投影单元，其包含了 T 个反投影

单元和一个上投影块，且每个反投影单元都包括一个上投影块和一个下投影块。各个投影单元之间采用密集连接的方式，由于密集连接方式是在维度层面拼接特征度，所以在每个上下投影块中都使用线性映射进行特征融合。此外为了充分发掘图像的深层特征信息，引入了残差学习的思想。如式(6)所示。

$$F_{DF,t}=H_{RBP,t}\left(\varphi\left(\psi\left(F_{DF,t-1},F_{DF,t-2},\cdots,F_{DF,1}\right),F_{IN,t-1}\right)\right) \tag{6}$$

其中 $H_{RBP,t}(\cdot)$, $\psi(\cdot,\cdot)$ 和 $\varphi(\cdot,\cdot)$ 分别表示第 t 个反投影单元、级联函数和逐元素求和函数， $F_{DF,t}$ ， $F_{DF,t\text{-}1}$ 和 $F_{DF,t\text{-}2}$ 分别代表第 t、t-1 和 t-2 个反投影单元输出的深层特征信息， $F_{IN,t-1}$ 表示第 t-1 个反投影单元输入的深层特征信息。即使用跳跃链接的方法，将第 t-1 个反投影单元输入的特征信息与级联了先前 t-1 个反投影单元的输出的特征信息的逐元素之和作为第 t 个反投影单元的输入。

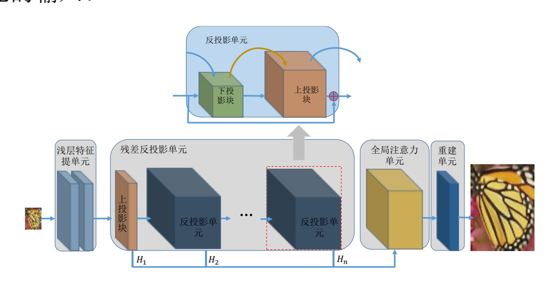


图 3 深度残差反投影注意力网络结构

Fig. 3　Depth residual backprojection attention network structure

每个反投影单元都是由一个上投影块和一个下投影块构成，其中上投影块和下投影块的结构分别如图 4、5 所示。

a) 上投影块。上投影块的输入是级联了此投影单元之前各投影单元中下投影块的输出，即第 t 个上投影块的输入是 $[L^1,\cdots,L^{t-1}]$ ，随后使用级联层将该投影单块的输入级联在一起，此外为了防止维度爆炸，采用了卷积核大小为 1*1 的卷积层降低特征图的维度从而获得特征 $L^{\tilde{t}-1}$,随后对 $L^{\tilde{t}-1}$ 进行上采样和下采样操作分别得到 H_0^t 和 L_0^t ，并计算 $L^{\tilde{t}-1}$ 和 L_0^t 之间的误差 e_t^l ,并使用误差 e_t^l 纠正 HR 特征和 LR 特征之间的映射关系。

b) 下投影块。下投影块的输入同样是级联了此投影单元之前各投影单元中上投影块的残差学习的结果，对输入的特征信息依次进行级联、线性映射得到特征图 $H^{\tilde{t}}$ ，随后依次进行下采样及上采样操作并计算重构误差 e_t^h ,并使用次重构误差指导重建 LR 特征图。

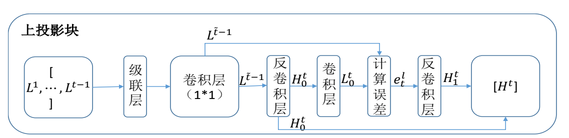


图 4 上投影块结构

Fig. 4　Up-projection block

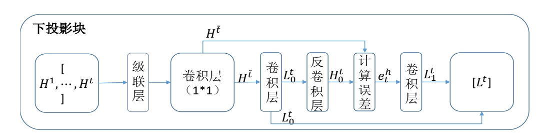


图 5 下投影块结构

Fig. 5　Down-projection block

2.3　全局注意力单元

在重建预测图像时，使用各投影单元中上投影块的输出结果。但是实际各个阶段学习到的特征信息具有个别差异性，并且相同阶段的不同通道之间特征信息也具有差别，导致在最后重建预测图像时其的贡献也各不相同。针对这些问题本文引入全局注意力机制，即为投影单元产生的 HR 特征信息的各通道都分配以不同的注意力，从而发掘更多的细节信息。

全局注意力单元的结构如图 6 所示，其中 ⊗ 表示逐元素相乘。该模块的输入是级联了各个反投影单元输出的 HR 特征图 $X=[H_1,\cdots,H_n]$ ，其包含了 $C=m\times n$ 个空间大小为 $H\times W$ 的特征图，即第 t 个反投影单元输出的 HR 特征图 H_t 的维度大小为 m。对于输入大小为 $H\times W\times C$ 的特征信息，本文使用全局池化操作进行逐通道的统计分析，其定义如式(7)所示。

$$z_c=H_{GP}(x)=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x_c(i,j) \tag{7}$$

其中 $x_c(i,j)$ 表示第 c 张特征图 (i,j) 处的特征信息， z_c 表示对第 c 张特征图的统计。然后对统计后的特征信息依次使用卷积层、ReLU 激活函数、卷积层对通道信息进行分析，最后使用门控函数计算各通道分配的注意力资源，并将计算得到的注意力分配值逐元素的和输入的特征图进行乘积操作。整个过程可以公式化为

$$\tilde{X}=f(W_U\delta(W_DZ))\cdot X \tag{8}$$

其中 $\delta(\cdot)$ 和 $f(\cdot)$ 分别表示 ReLU 激活函数和 Sigmoid 门控函数， W_U 和 W_D 分别代表相应卷积层的权重矩阵， X 和 $\tilde{X}$ 分别表示大小为 $H\times W\times C$ 输入的特征信息和分配过注意力资源的特征信息, z 表示平均池化而得的全局特征信息。

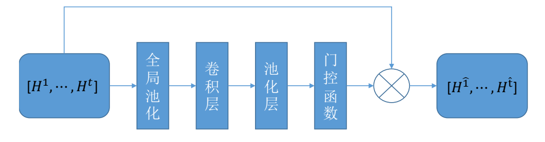


图 6 全局注意力结构

Fig. 6　Schematic diagram of global attention unit

3　实验

本文使用 DIV2K[13]数据集中的 800 张图片训练所提出的网络结构，此外选用四个公开的标准数据集测试本文的算法，选用的四个标准数据集分别是：Set5[14]、Set14[15]、BSD100[16]、Urban100[17]，并使用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性(Structural SIMilarity index, SSIM)[18]的取值作为对预测结果的客观评判标准。在实验中对于不同的上采样尺度，采用了不同大小的卷积核进行反卷积操作。其中在缩放因子为 2 时，采用了 stride = 2 ，padding = 2 和卷积核大小为 6×6 的卷积层，在缩放因子为 4 时 stride = 4 ，padding = 2 卷积核核大小为 8×8 ，缩放因子为 8 时 stride = 8 ，padding = 2 卷积核核大小为 12×12 。此外初始化学习率为 1E-4，并且在迭代 10^6 过程中每迭代 5×10^5 次学习率减少 10 倍。另外使用 L_1 范数损失函数和动量为 0.9 的 Adam[19]优化器，优化本文算法。所有的实验均部署在 Nvidia TITAN X (Pascal) GPU 以及 Intel(R) Xeon(R) W-2125 CPU 环境下。

3.1　范数损失函数和 MSE 损失函数结果对比

损失函数代表着网络模型的学习目标，所以损失函数的选择通常对实验结果具有很大影响。在 SISR 任务中通常使用(mean squared error, MSE)损失函数和 范数损失函数，其中 MSE 损失函数更倾向于获得更高的峰值信噪比(PSNR)，但是近些年研究者[20]发现使用 L_1 范数可以加速网络模型的收敛。为了选择出更适合本文算法的损失函数，本文分别使用 MSE 损失函数和 L_1 范数损失函数构造模型 I 和模型 II 。为了保证对比实验结果的公平性，模型 I 和模型 II 使用相同的网络深度和相同大小的卷积核，以及其余设置也都相同(缩放因子为 4，反投影单元为 6)。将使用相同训练集获得的网络模型 I 和模型 II ，分别在 Set5 和 Set14 数据集上进行测试，测试结果如图 7、8 所示。

从图 7、8 可以发现在 Set5 数据集上，使用 L_1 范数损失函数作为优化目标的平均 PSNR 和 SSIM 比使用 MSE 损失函数作为优化目标的平均 PSNR 和 SSIM 高 0.10dB,0.0007。在 Set14 数据集上，使用 L_1 范数损失函数作为优化目标的平均 PSNR 和 SSIM 比使用 MSE 损失函数作为优化目标的平

均 PSNR 和 SSIM 高 0.73dB 和 0.0227。综上所述，本文更适合选用 L_1 范数损失函数作为优化目标。

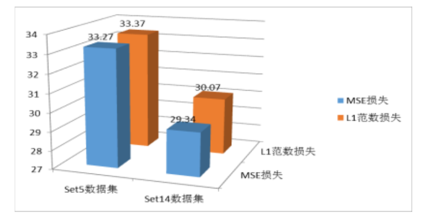


图 7　范数损失和 MSE 损失在不同数据集上的 PSNR

Fig. 7　Comparison of SSIM between L_1 -norm loss and MSE loss under different datasets

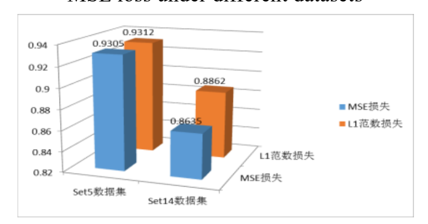


图 8　范数损失和 MSE 损失在不同数据集上的 SSIM

Fig. 8　Comparison of SSIM between L_1 -norm loss and MSE loss under different datasets

3.2　模型分析

本文提出的深度残差反投影注意力网络不仅融合了残差思想，而且结合了注意力机制以提高预测图像的质量。为了验证网络结构中每部分的有效性，本文通过依次删除残差思想和注意力机制的方式，构造了具有相同网络深度和相同大小的卷积核的不同模型分别记为模型Ⅲ、模型Ⅳ和模型Ⅴ。其中模型Ⅲ是文献[11]中提出的 D-DBPN 网络结构，模型Ⅳ是本文提出的只融合了全局注意力机制的模型，模型Ⅴ既是本文提出的融合了残差思想和全局注意力机制的网络模型。模型Ⅲ、模型Ⅳ和模型Ⅴ均是使用了 6 个反投影块，且当缩放因子为2､4和8时，卷积核大小分别为 6×6 、8×8 和 12×12 ，其实验结果如表 1 所示。

从表 1 中，本文可以发现模型Ⅳ和模型Ⅴ相对于模型Ⅲ来说，预测图像在 PSNR 和 SSIM 两个指标上均有较大的提升。当缩放因子为 4 时，模型Ⅳ和模型Ⅴ在 Set5 数据集上的 PSNR 和 SSIM 相对模型Ⅲ分别提高了 1.46dB,0.039 和 1.38dB,0.038，在 Set14 数据集上分别提高了 0.82dB,0.087 和 1.55dB,0.108。当缩放因子为 8 时，模型Ⅳ和模型Ⅴ在 Set5 数据集上的 PSNR 和 SSIM 相对模型Ⅲ分别提高了 0.79dB,0.076 和 1.27dB,0.081，在 Set14 数据集上分别提高了 1.09dB,0.145 和 1.13dB,0.146。另外本文还可以发现只有在 Set5 数据集上且缩放因子为 4 时，模型Ⅴ预测图像的质量低于模型Ⅳ。但是其余的实验结果均表明模型Ⅴ的实验结果优于模型Ⅳ的实验结果，所以本文认为模型Ⅴ总体上要优于模型Ⅳ。

表 1　分别在 Set5 和 Set14 数据集上对比模型Ⅰ、Ⅱ和Ⅲ

Tab. 1　Performance of model Ⅰ, Ⅱ and Ⅲ on the Set5 and Set14 datasets

采样倍数	残差学习	全局注意力机制	Set5 psnr	Set5 ssim	Set4 psnr	Set4 ssim
	×	×	31.99	0.893	28.52	0.778
X4	×	√	**33.45**	**0.932**	29.34	0.865
	√	√	33.37	0.931	**30.07**	**0.886**
	×	×	26.86	0.773	24.92	0.638
X8	×	√	27.65	0.849	26.01	0.783
	√	√	**28.13**	**0.854**	**26.05**	**0.784**

3.3　参数规模对比

为了检验本文算法参数规模的大小，将本文算法的参数规模和部分主流算法的参数规模进行对比实验，选取的部分主流算法包括 SRCNN[1],VDSR[2],FSRCNN[8],LapSRN[10], D-DBPN[11],EDSR[20], MemNet[21], RCAN[22]，实验结果如图 9、10 所示。其中图 9、10 分别是当缩放因子为 4 和 8 时，本文算法和部分主流算法在 Set5 数据集上测试时的参数规模对比结果。

从图中本文可以发现在缩放因子为 4 时，本文算法在 PSNR 指标上高出次优算法 0.81dB 的情况下，参数规模仍然维持在小于次优算法的参数规模。当缩放因子为 8 时，本文算法的参数规模有所增加，但是 PSNR 指标仍然高于次优算法 0.88dB。总体来说，本文算法在 PSNR 指标上获得比较优异的表现的情况下，参数规模仍然维持在较客观的水平上。

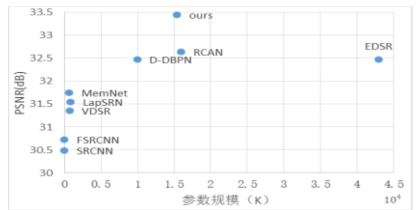


图 9　set5 数据集下各主流算法的 x4 模型的参数规模对比

Fig. 9　Parameter scale comparison of x4 model of mainstream algorithms under set5 dataset

图 10　set5 数据集下各主流算法的 x8 模型的参数规模对比

Fig. 10　Parameter scale comparison of x8 model of mainstream algorithms under set5 dataset

3.4　对比先进算法

为了验证本文算法的有效性，本节分别对本文算法和部分主流算法在不同数据集上的预测结果进行对比分析。实验采用的四个公开数据集分别是由比利埃大学法国贝尔实验室发布的 Set5 和 Set14，伯克利大学发布的 BSDS100 以及最近有黄等人提供的 Urban100 数据集，它们包含了不同场景的自然图像。对比的部分主流算法包括 SRCNN[1], VDSR[2], FSRCNN[11], LapSRN[10], D-DBPN[11], EDSR[20], MemNet[21], RCAN[22] SCN [23], SRMDNF[24] 和 RDN[25]等，分别对比这些主流算法在不同数据集上缩放因子取值分别为 2、4 和 8 时的峰值信噪比(PSNR)、结构相似形(SSIM)以及主观的视觉效果。此外实验使用公开的代码，并使用相同的训练集重新进行实验，以保证实验的公平性原则。

实验结果分别如表 2~4 所示，其分别是缩放因子为 2、4 和 8 时，部分主流算法在不同数据集上的 PSNR 和 SSIM 取值。从表 2 中可以发现在缩放因子为 2 时，本文算法在 Set5 数据集上未取得最优的 PSNR，但是在其余数据集上均取得了最优的 PSNR，其中在 BSDS100 数据集上相对于其他主流算法提升的最明显，相对于 RCAN 提高了 2.93dB，此外在 SSIM 指标上本文算法均达到了最优值。观察表 3、4 本文可以发现在缩放因子为 4 和 8 时，本文算法无论是在 PSNR 指标和 SSIM 指标上均取得了最优的表现，其中也是在 BSDS100 数据集上，取得了较大的提升。故此，本文算法在总体上取得了优异的表现，并且缩放因子越大，优势越明显。

本文从数据集 Set5、Set14、BSDS100 和 Urban100 中分别选取了“women”,“comic”,“119082”,“img_037”等图像，对比分析本文算法和部分主流算法在缩放因子为 4 时的重建结果的视觉体验。此外为了清晰的观察到细节信息的对比结果，本文对重建得到的结果进行部分区域放大，其实验

结果分别如图 11~14 所示。从图中可以发现 Bicubic 插值法重建的图像几乎观察不到轮廓等细节信息，VDSR 和 LapSRN 算法重建的结果虽然得到了些许的提升，但是仍然缺少了部分的细节信息。EDSR、D-DBPN 和 RCAN 等算法重建的结果获得了较好的视觉体验，但是相对于本文算法依然缺少锐利的边缘信息。综上无论在客观指标上，还是在主观视觉体验上本文算法都取得了优异的表现。

表 2　各 SISR 算法的 x2 模型在不同数据集上的表现

Tab. 2　Average performance of x2 models of various SISR algorithm on different data sets

算法	Set5		Set14		BSDS100		Urban100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403
SRCNN	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946
FSRCNN	37.05	0.9560	32.66	0.9090	31.53	0.8920	29.88	0.9020
VDSR	37.53	0.9590	33.05	0.9130	31.90	0.8960	30.77	0.9140
LapSRN	37.52	0.9591	33.08	0.9130	31.08	0.8950	30.41	0.9101
MemNet	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195
EDSR	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351
SRMDNF	37.79	0.9601	33.32	0.9159	32.05	0.8985	31.33	0.9204
D-DBPN	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324
RDN	38.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353
RCAN	**38.27**	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384
本文算法	38.12	**0.9663**	**34.52**	**0.9329**	**35.34**	**0.9595**	**33.38**	**0.9565**

表 3　各 SISR 算法的 x4 模型在不同数据集上的表现

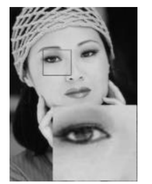
Tab. 3　Average performance of x4 models of various SISR algorithm on different data sets

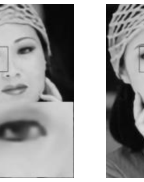
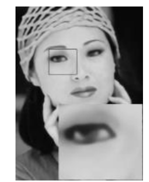
算法	Set5		Set14		BSDS100		Urban100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577
SRCNN	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221
FSRCNN	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280
VDSR	31.35	0.8830	28.02	0.7680	27.29	0.0726	25.18	0.7540
LapSRN	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560
MemNet	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630
EDSR	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033
SRMDNF	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731
D-DBPN	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946
RDN	33.36	0.9313	28.81	0.7871	27.72	0.7419	26.61	0.8028
RCAN	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087
本文算法	**33.44**	**0.9319**	**29.35**	**0.8644**	**30.92**	**0.8944**	**27.49**	**0.8777**

表 4　各 SISR 算法的 x8 模型在不同数据集上的表现

Tab. 4　Average performance of x8 models of various SISR algorithm on different data sets

算法	Set5		Set14		BSDS100		Urban100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	24.40	0.6580	23.10	0.5660	23.67	0.5480	20.74	0.5160
SRCNN	25.33	0.6900	23.76	0.5910	24.13	0.5660	21.29	0.5440
FSRCNN	20.13	0.5520	19.75	0.4280	24.21	0.5680	21.32	0.5380
SCN	25.59	0.7071	24.02	0.6025	24.30	0.5698	21.52	0.5571
VDSR	25.93	0.7240	24.26	0.6140	24.49	0.5830	21.70	0.5710
LapSRN	26.15	0.7380	24.35	0.6200	24.54	0.5860	21.81	0.5810
MemNet	26.16	0.7414	24.38	0.6199	24.58	0.5842	21.89	0.5825
EDSR	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221
D-DBPN	27.21	0.7840	25.13	0.6480	24.88	0.6010	22.73	0.6312
RCAN	27.31	0.7878	25.23	0.6511	24.98	0.6058	23.00	0.6452
本文算法	**28.19**	**0.8555**	**26.09**	**0.7840**	**27.64**	**0.8248**	**23.58**	**0.7805**



(a)HR　(b)Bicubic　(c)VDSR　(d)LapSRN

(e)EDSR　(f)D-DBPN　(g)RCAN　(h)本文

图 11　Set5 数据集 women 重建结果对比

Fig. 11　Comparison of reconstruction results of the women in the Set5 dataset





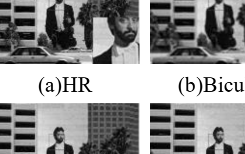
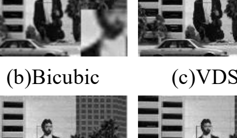

(a)HR　(b)Bicubic　(c)VDSR　(d)LapSRN




(e)EDSR　(f)D-DBPN　(g)RCAN　(h)本文

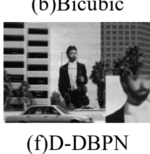
图 12　Set14 数据集 comic 重建结果对比

Fig. 12　Comparison of reconstruction results of the comic in Set14 dataset



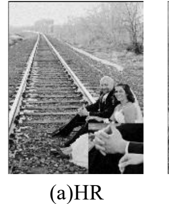
(a)HR　(b)Bicubic　(c)VDSR　(d)LapSRN



(e)EDSR　(f)D-DBPN　(g)RCAN　(h)本文

图 13　BSDS100 数据集 119082 重建结果对比

Fig. 13　Comparison of reconstruction results of the 119082 in the BSDS100 dataset






(a)HR　(b)Bicubic　(c)VDSR　(d)LapSRN



(e)EDSR　(f)D-DBPN　(g)RCAN　(h)本文

图 14　Urban100 数据集 img_037 重建结果对比

Fig. 14　Comparison of reconstruction results of the img_037 of in Urban100 dataset

4　结束语

本文提出了深度残差反投影全局注意力网络，其主要使

用迭代反投影的方法并融合了残差学习和全局注意力机制，缓解了学习过程中特征信息利用不充分和高频信息丢失等问题，同时充分利用了特征图之间的差异性，使得重建预测图像时发掘更有用的高频信息。实验结果证明了，本文算法在 PSNR 指标和 SSIM 指标上的优越性，同时得到的预测图像细节信息更加丰富，视觉体验也更好。

参考文献：

[1] Dong C, LOY C C, HE K, *et al.* Image super-resolution using deep convolutional networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38 (2): 295-307.

[2] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]// Proceedings of the 2016 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) . Piscataway: IEEE, 2016. 1646-1654. [doi: 10. 1109/CVPR. 2016. 182]

[3] 王晓明, 黄凤, 刘少鹏, 徐涛. 改进的单幅图像自学习超分辨率重建方法 [J]. 计算机应用研究, 2019, 36 (08): 2534-2538. (Wang Xiaoming, Huang Feng, Liu Shaopeng, Xu Tao. Improved single-image self-learning super-resolution reconstruction method [J]. Computer Application Research, 2019, 36 (08): 2534-2538.)

[4] 邵文泽, 韦志辉. 基于各向异性 MRF 建模的多帧图像变分超分辨率重建 [J]. 电子学报, 2009, 37 (6): 1256-1263. (Shao Wenze, Wei Zhihui. Variational super-resolution reconstruction of multi-frame images based on anisotropic MRF modeling [J]. Chinese Journal of Electronics, 2009, 37 (6): 1256-1263.)

[5] 李展, 张庆丰, 孟小华, 等. 多分辨率图像序列的超分辨率重建 [J]. 自动化学报, 2012, 38 (11): 1804-1814. (Li Zhan, Zhang QingFeng, Meng Xiaohua, *et al.* Super-resolution reconstruction of multi-resolution image sequences [J]. Acta Automatica Sinica, 2012, 38 (11): 1804-1814.)

[6] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network [C]// Proc of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . Honolulu, HI, USA: IEEE, 2017: 2790-2798

[7] Kim J, Kwon L J, MU L K. Deeply-recursive convolutional network for image super-resolution [C]// Proceedings of the 2016 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 1637 – 1645. [doi: 10. 1109/CVPR. 2016. 181]

[8] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural networ k [C]// Proceedings of the 2016 European Conference on Computer Vision. Amsterdam, Netherlands: Springer Verlag, 2016: 391-407

[9] Shi W, Caballero J, HUSZÁR F, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1874-1883.

[10] Lai W S, Huang J B, Ahuja N, *et al.* Deep laplacian pyramid networks for fast and accurate super-resolution [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . Piscataway: IEEE, 2017: 624-632

[11] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution [C]// Proceeding of the 2018 IEEE International Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 1664-1673

[12] He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. [C]// Proceeding of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . Las Vegas, NV, USA: IEEE, 2016: 770-778

[13] Timofte R, Agustsson E, Van G L, *et al.* NTIRE 2017 challenge on single image super-resolution: Methods and results [C]// Proceeding of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2017: 1110-1121.

[14] Bevilacqua M, Roumy A, Guillemot C, *et al.* Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C]// Proceeding of the 2012 British Machine Vision Conference. Durham: BMVA Press, 2012: 135. 1 – 135. 10

[15] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations [C]// Proceeding of the 7th International Conference on Curves and Surfaces. Berlin: Springer, 2010: 711-730

[16] Martin D, Fowlkes C, Tal D, *et al.* A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C]// proceeding of the 2001 Eighth IEEE International Conference on Computer Vision. Canada: IEEE 2001: 416-423

[17] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars [C]// Proceedings of the 2015 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 5197-5206.

[18] YE Y X, Shan J, Bruzzone L, et a1. Robust registration of multimodal remote sensing images based on structural similarity [J]. IEEE Transactions on Geoseience and Remote Sensing, 2017, 55 (5): 2941—2958.

[19] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. [2014-12-22]. https://arxiv. org/abs/1412. 6980.

[20] Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks for single image super-resolution [C]// CVPRW 2017: Proceeding of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington, DC: IEEE Computer Society, 2017: 1132 – 1140.

[21] Tai Y. , Yang J, Liu X, *et al.* : Memnet: a persistent memory network for image restoration [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Venice, Italy: IEEE 2 017: 4539-4547.

[22] Zhang Y, Li K, Li K, *et al.* Image super-resolution using very deep residual channel attention networks [C]// Proceedings of the 2018 European Conference on Computer Vision. Munich Germany: Springer Verlag 2018: 294-310.

[23] Wang Z, Liu D, Yang J, *et al.* Deep networks for image super-resolution with sparse prior [C]// Proceeding of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE 2015: 370-378.

[24] Zhang K, ZUO W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations [C]// Proceeding of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE 2018: 3262-3271

[25] Zhang Y, Tian Y, Kong Y, *et al.* Residual dense network for image super-resolution [C]// proceeding of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Athens, Greece: IEEE 2018: 2472-2481.